

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Life Technologies, Inc.
8717 Grovemont Circle
Gaithersburg, MD 20884-9980
APPLICANTS/INVENTORS: Chatterjee, Deb K.

(ii) TITLE OF INVENTION: Cloned DNA Polymerases from *Thermotoga*
and Mutants Thereof

(iii) NUMBER OF SEQUENCES: 22

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
(B) STREET: 1100 New York Ave., N.W., Suite 600
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20005

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: To be assigned
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/525,057
(B) FILING DATE: 08-SEP-1995
(C) CLASSIFICATION:

(viii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/537,397
(B) FILING DATE: 02-OCT-1995
(C) CLASSIFICATION:

(ix) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/537,400
(B) FILING DATE: 02-OCT-1995
(C) CLASSIFICATION:

- (vii) PRIOR APPLICATION DATA:
 (A) APPLICATION NUMBER: 08/576,759
 (B) FILING DATE: 21-DEC-1995
 (C) CLASSIFICATION:

- (vii) PRIOR APPLICATION DATA:
 (A) APPLICATION NUMBER: To be assigned
 (B) FILING DATE: 14-AUG-1996
 (C) CLASSIFICATION:

- (viii) ATTORNEY/AGENT INFORMATION:
 (A) NAME: Esmond, Robert W.
 (B) REGISTRATION NUMBER: 32,893
 (C) REFERENCE/DOCKET NUMBER: 0942.280PC03

- (ix) TELECOMMUNICATION INFORMATION:
 (A) TELEPHONE: 202-371-2600
 (B) TELEFAX: 202-371-2540

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAGCTCACGG GGGATGCAGG AAA

23

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2682 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

-64-

CGAGAC	TATTTCTCTT	TGATGGCACA	GCCCTGGCCT	ACAGGGCATA	TTACGCCCTC
ATCCC	TTCCACATC	CACAGGAATT	CCAACGAACG	CCGTCTATGG	CGTTGCCAGG
CGTTA	AATTCATTAA	GGAACACATT	ATACCCGAAA	AGGACTACGC	GGCTGTGGCC
CAAGA	AGGCAGCGAC	GTTTCAGACAC	AAACTGCTCG	TAAGCGACAA	GGCGCAAAGG
GACTC	CGGCTCTTCT	AGTTCAGCAG	CTACCTTACA	TCAAGCGGCT	GATAGAAGCT
TTTCA	AAGTGCTGGA	GCTGGAGGGA	TACGAAGCAG	ACGATATCAT	CGCCACGCTT
CAGGG	CTGCACGTTT	TTTGATGAGA	TTTTCATTA	TAACCGGTGA	CAAGGATATG
ACTTG	TAAACGAGAA	GATAAAGGTC	TGGAGAATCG	TCAAGGGGAT	ATCGGATCTT
TTACG	ATTCGAAAAA	GGTGAAAGAA	AGATACGGTG	TGGAACCACA	TCAGATACCG
TCTAG	CACTGACGGG	AGACGACATA	GACAACATTC	CCGGTGTAAC	GGGAATAGGT
GACCG	CTGTACAGCT	TCTCGGCAAG	TATAGAAATC	TTGAATACAT	TCTGGAGCAT
TGAAC	TCCCCCAGAG	AGTGAGAAAG	GCTCTCTTGA	GAGACAGGGA	AGTTGCCATC
TAAAA	AACTTGCAAC	TCTGGTGACG	AACGCACCTG	TTGAAGTGGA	CTGGGAAGAG
ATACA	GAGGATACGA	CAAGAGAAAA	CTACTTCCGA	TATTGAAAGA	ACTGGAGTTT
CATCA	TGAAGGAACT	TCAACTGTAC	GAAGAAGCAG	AACCCACCGG	ATACGAAATC
GGATC	ATAAGACCTT	CGAAGATCTC	ATCGAAAAGC	TGAAGGAGGT	TCCATCTTTT
GGACC	TTGAAACGTC	CTCCTTGGAC	CCGTTCAACT	GTGAGATAGT	CGGCATCTCC
GTTCA	AACCGAAAAC	AGCTTATTAC	ATTCCACTTC	ATCACAGAAA	CGCCACAAT
TGAAA	CACTGGTGCT	GTCGAGATTG	AAAGAGATCC	TCGAAGACCC	GTCTTCGAAG
GGGTC	AGAACCTGAA	GTACGACTAC	AAGGTTCTTA	TGGTAAAGGG	TATATCGCCA
TCCGC	ATTTTGACAC	GATGATAGCT	GCATATTTGC	TGGAGCCAAA	CGAGAAAAAA
TCTCG	AAGATCTGTC	TTTGAAATTT	CTCGGATACA	AAATGACGTC	TTATCAGGAA
GTCGT	TTTCCTCACC	ACTTTTTTGGT	TTCAGCTTTG	CGGATGTTCC	GGTAGACAAG
CGAAT	ACTCCTGCGA	GGATGCAGAC	ATCACTTATA	GGCTCTACAA	GATACTCAGC
GCTCC	ATGAAGCGGA	ACTTGAGAAC	GTCTTCTACA	GGATAGAGAT	GCCGTTGGTG
CTTGG	CACGAATGGA	ATTCAACTGG	GTGTATGTTG	ACACAGAATT	CCTGAAAAAG
GGAGG	AGTACGGCAA	AAAGCTCGAG	GAAGTGGCCG	AAAAAATCTA	CCAGATAGCA

GGTGAGCCCT TCAACATCAA TTCTCCAAAA CAGGTTTCAA ACATCCTTTT TGAGAAGCTG 1680
 GGAATAAAAC CCCGTGGAAA AACGACAAAA ACAGGAGATT ACTCTACCAG GATAGAGGTG 1740
 TTGGAAGAGA TAGCGAATGA GCACGAGATA GTACCCCTCA TTCTCGAGTT CAGAAAGATC 1800
 CTGAAACTGA AATCGACCTA CATAGACACC CTTCCGAAAC TTGTGAACCC GAAAACCGGA 1860
 AGATTTTCATG CATCTTTCCA CCAGACGGGT ACCGCCACTG GCAGGTTGAG TAGCAGTGAT 1920
 CCAAATCTTC AGAATCTTCC GACAAAGAGC GAAGAGGGAA AAGAAATTAG AAAAGCGATT 1980
 GTGCCCCAGG ATCCAGACTG GTGGATCGTC AGTGCGGATT ATTCCCAAAT AGAACTCAGA 2040
 ATCCTCGCTC ATCTCAGTGG TGATGAGAAC CTTGTGAAGG CCTTCGAGGA GGGCATCGAT 2100
 GTGCACACCT TGA CTGCCTC CAGGATCTAC AACGTAAAGC CAGAAGAAGT GAACGAAGAA 2160
 ATGCGACGGG TTGGAAGAT GGTGAACTTC TCTATAATAT ACGGTGTCAC ACCGTACGGT 2220
 CTTTCTGTGA GACTTGGAAT ACCGGTTAAA GAAGCAGAAA AGATGATTAT CAGCTATTTT 2280
 ACACTGTATC CAAAGGTGCG AAGCTACATC CAGCAGGTTG TTGCAGAGGC AAAAGAGAAG 2340
 GGCTACGTCA GGA CTCTCTT TGGAAGAAAA AAGATATTTC CCCAGCTCAT GGCAAGGGAC 2400
 AAGAACACCC AGTCCGAAGG CGAAAGAATC GCGATAAACA CCCCATTCA GGGAAGTGGC 2460
 GCAGATATAA TAAAATTGGC TATGATAGAT ATAGACGAGG AGCTGAGAAA AAGAAACATG 2520
 AAATCCAGAA TGATCATTCA GGTTCATGAC GAACTGGTCT TCGAGGTTCC CGATGAGGAA 2580
 AAAGAAGAAC TAGTTGATCT GGTGAAGAAC AAAATGACAA ATGTGGTGAA ACTCTCTGTG 2640
 CCTCTTGAGG TTGACATAAG CATCGGAAAA AGCTGGTCTT GA 2682

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 893 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Ala Arg Leu Phe Leu Phe Asp Gly Thr Ala Leu Ala Tyr Arg Ala
 1 5 10 15

Tyr Tyr Ala Leu Asp Arg Ser Leu Ser Thr Ser Thr Gly Ile Pro Thr
 20 25 30
 Asn Ala Val Tyr Gly Val Ala Arg Met Leu Val Lys Phe Ile Lys Glu
 35 40 45
 His Ile Ile Pro Glu Lys Asp Tyr Ala Ala Val Ala Phe Asp Lys Lys
 50 55 60
 Ala Ala Thr Phe Arg His Lys Leu Leu Val Ser Asp Lys Ala Gln Arg
 65 70 75 80
 Pro Lys Thr Pro Ala Leu Leu Val Gln Gln Leu Pro Tyr Ile Lys Arg
 85 90 95
 Leu Ile Glu Ala Leu Gly Phe Lys Val Leu Glu Leu Glu Gly Tyr Glu
 100 105 110
 Ala Asp Asp Ile Ile Ala Thr Leu Ala Val Arg Ala Ala Arg Phe Leu
 115 120 125
 Met Arg Phe Ser Leu Ile Thr Gly Asp Lys Asp Met Leu Gln Leu Val
 130 135 140
 Asn Glu Lys Ile Lys Val Trp Arg Ile Val Lys Gly Ile Ser Asp Leu
 145 150 155 160
 Glu Leu Tyr Asp Ser Lys Lys Val Lys Glu Arg Tyr Gly Val Glu Pro
 165 170 175
 His Gln Ile Pro Asp Leu Leu Ala Leu Thr Gly Asp Asp Ile Asp Asn
 180 185 190
 Ile Pro Gly Val Thr Gly Ile Gly Glu Lys Thr Ala Val Gln Leu Leu
 195 200 205
 Gly Lys Tyr Arg Asn Leu Glu Tyr Ile Leu Glu His Ala Arg Glu Leu
 210 215 220
 Pro Gln Arg Val Arg Lys Ala Leu Leu Arg Asp Arg Glu Val Ala Ile
 225 230 235 240
 Leu Ser Lys Lys Leu Ala Thr Leu Val Thr Asn Ala Pro Val Glu Val
 245 250 255
 Asp Trp Glu Glu Met Lys Tyr Arg Gly Tyr Asp Lys Arg Lys Leu Leu
 260 265 270
 Pro Ile Leu Lys Glu Leu Glu Phe Ala Ser Ile Met Lys Glu Leu Gln
 275 280 285
 Leu Tyr Glu Glu Ala Glu Pro Thr Gly Tyr Glu Ile Val Lys Asp His
 290 295 300

657094360

Lys Thr Phe Glu Asp Leu Ile Glu Lys Leu Lys Glu Val Pro Ser Phe
 305 310 315 320
 Ala Leu Asp Leu Glu Thr Ser Ser Leu Asp Pro Phe Asn Cys Glu Ile
 325 330 335
 Val Gly Ile Ser Val Ser Phe Lys Pro Lys Thr Ala Tyr Tyr Ile Pro
 340 345 350
 Leu His His Arg Asn Ala His Asn Leu Asp Glu Thr Leu Val Leu Ser
 355 360 365
 Lys Leu Lys Glu Ile Leu Glu Asp Pro Ser Ser Lys Ile Val Gly Gln
 370 375 380
 Asn Leu Lys Tyr Asp Tyr Lys Val Leu Met Val Lys Gly Ile Ser Pro
 385 390 395 400
 Val Tyr Pro His Phe Asp Thr Met Ile Ala Ala Tyr Leu Leu Glu Pro
 405 410 415
 Asn Glu Lys Lys Phe Asn Leu Glu Asp Leu Ser Leu Lys Phe Leu Gly
 420 425 430
 Tyr Lys Met Thr Ser Tyr Gln Glu Leu Met Ser Phe Ser Ser Pro Leu
 435 440 445
 Phe Gly Phe Ser Phe Ala Asp Val Pro Val Asp Lys Ala Ala Glu Tyr
 450 455 460
 Ser Cys Glu Asp Ala Asp Ile Thr Tyr Arg Leu Tyr Lys Ile Leu Ser
 465 470 475 480
 Met Lys Leu His Glu Ala Glu Leu Glu Asn Val Phe Tyr Arg Ile Glu
 485 490 495
 Met Pro Leu Val Asn Val Leu Ala Arg Met Glu Phe Asn Trp Val Tyr
 500 505 510
 Val Asp Thr Glu Phe Leu Lys Lys Leu Ser Glu Glu Tyr Gly Lys Lys
 515 520 525
 Leu Glu Glu Leu Ala Glu Lys Ile Tyr Gln Ile Ala Gly Glu Pro Phe
 530 535 540
 Asn Ile Asn Ser Pro Lys Gln Val Ser Asn Ile Leu Phe Glu Lys Leu
 545 550 555 560
 Gly Ile Lys Pro Arg Gly Lys Thr Thr Lys Thr Gly Asp Tyr Ser Thr
 565 570 575
 Arg Ile Glu Val Leu Glu Glu Ile Ala Asn Glu His Glu Ile Val Pro
 580 585 590

55310" 55310" 55310"

Leu	Ile	Leu	Glu	Phe	Arg	Lys	Ile	Leu	Lys	Leu	Lys	Ser	Thr	Tyr	Ile
		595					600					605			
Asp	Thr	Leu	Pro	Lys	Leu	Val	Asn	Pro	Lys	Thr	Gly	Arg	Phe	His	Ala
	610					615					620				
Ser	Phe	His	Gln	Thr	Gly	Thr	Ala	Thr	Gly	Arg	Leu	Ser	Ser	Ser	Asp
625					630					635					640
Pro	Asn	Leu	Gln	Asn	Leu	Pro	Thr	Lys	Ser	Glu	Glu	Gly	Lys	Glu	Ile
				645					650					655	
Arg	Lys	Ala	Ile	Val	Pro	Gln	Asp	Pro	Asp	Trp	Trp	Ile	Val	Ser	Ala
			660					665					670		
Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Ile	Leu	Ala	His	Leu	Ser	Gly	Asp
		675					680					685			
Glu	Asn	Leu	Val	Lys	Ala	Phe	Glu	Glu	Gly	Ile	Asp	Val	His	Thr	Leu
	690					695					700				
Thr	Ala	Ser	Arg	Ile	Tyr	Asn	Val	Lys	Pro	Glu	Glu	Val	Asn	Glu	Glu
705					710					715					720
Met	Arg	Arg	Val	Gly	Lys	Met	Val	Asn	Phe	Ser	Ile	Ile	Tyr	Gly	Val
				725					730					735	
Thr	Pro	Tyr	Gly	Leu	Ser	Val	Arg	Leu	Gly	Ile	Pro	Val	Lys	Glu	Ala
			740					745					750		
Glu	Lys	Met	Ile	Ile	Ser	Tyr	Phe	Thr	Leu	Tyr	Pro	Lys	Val	Arg	Ser
		755					760					765			
Tyr	Ile	Gln	Gln	Val	Val	Ala	Glu	Ala	Lys	Glu	Lys	Gly	Tyr	Val	Arg
	770					775					780				
Thr	Leu	Phe	Gly	Arg	Lys	Arg	Asp	Ile	Pro	Gln	Leu	Met	Ala	Arg	Asp
785					790					795					800
Lys	Asn	Thr	Gln	Ser	Glu	Gly	Glu	Arg	Ile	Ala	Ile	Asn	Thr	Pro	Ile
				805					810					815	
Gln	Gly	Thr	Ala	Ala	Asp	Ile	Ile	Lys	Leu	Ala	Met	Ile	Asp	Ile	Asp
			820					825					830		
Glu	Glu	Leu	Arg	Lys	Arg	Asn	Met	Lys	Ser	Arg	Met	Ile	Ile	Gln	Val
		835					840					845			
His	Asp	Glu	Leu	Val	Phe	Glu	Val	Pro	Asp	Glu	Glu	Lys	Glu	Glu	Leu
	850					855					860				
Val	Asp	Leu	Val	Lys	Asn	Lys	Met	Thr	Asn	Val	Val	Lys	Leu	Ser	Val
865					870					875					880

Pro Leu Glu Val Asp Ile Ser Ile Gly Lys Ser Trp Ser
885 890

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Pro Ser Phe Ala Leu Asp Leu Glu Thr Ser Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Pro Val Phe Ala Phe Asp Thr Glu Thr Asp Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

1

5

10

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: peptide

Met Ile Val Ser Asp Ile Glu Ala Asn Ala
1 5 10

1

5

10

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: cDNA

GACGTTTCAA GCGCTAGGGC AAAAGA

(2) INFORMATION FOR SEQ ID NO:9:

(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (ii) / MOLECULE TYPE: peptide

Lys Met Val Asn Phe Ser Ile Ile Tyr Gly

1 5 10

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Lys Ala Ile Asn Phe Gly Leu Ile Tyr Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Lys Ala Ile Thr Phe Gly Ile Leu Tyr Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Lys Thr Phe Ile Tyr Gly Phe Leu Tyr Gly

5' TTT GAT GAG GAG GAG GAG GAG GAG GAG GAG GAG 3'

5

20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE/TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTATATTATA GAGTAGTTAA CCATCTTTCC A

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTAGGCCAGG GGCTGTGCCG GCAAAGAGAA ATAGTC

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAAGGATATC CTTGGCGCCG GTTATTATGA AAATC

35

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1310 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ATGGCGAGAC TATTTCTCTT TGATGGCACA GCCCTGGCCT ACAGGGCATA TTACGCCCTC	60
GACAGATCCC TTTCCACATC CACAGGAATT CCAACGAACG CCGTCTATGG CGTTGCCAGG	120
ATGCTCGTTA AATTCATTAA GGAACACATT ATACCCGAAA AGGACTACGC GGCTGTGGCC	180
TTGACAAGA AGGCAGCGAC GTTCAGACAC AACTGCTCG TAAGCGACAA GGCGCAAAGG	240
CCAAAGACGC CGGCTCTTCT AGTTCAGCAG CTACCTTACA TCAAGCGGCT GATAGAAGCT	300
CTTGGTTTCA AAGTGCTGGA GCTGGAAGGG TACGAAGCAG ACGATATCAT CGCCACGCTT	360
GCAGCAAAGG GCTGCGCGTT TTTTGATGAG ATTTTCATAA TAACCGGTGA CAAGGATATG	420
CTTCAACTTG TAAACGAGAA GATAAAGGTC TGGAGAATCG TCAAGGGGAT ATCGGATCTT	480
GAGCTTTACG ATTCGAAAAA GGTGAAAGAA AGATACGGTG TGGAACCACA TCAGATACCG	540
GATCTTCTAG CACTGACGGG AGACGACATA GACAACATTC CCGGTGTAAC GGAATAGGT	600
GAAAAGACCG CTGTACAGCT TCTCGGCAAG TATAGAAATC TTGAATACAT TCTGGAGCAT	660
GCCCGTGAAC TCCCCCAGAG AGTGAGAAAG GCTCTCTTGA GAGACAGGGA AGTTGCCATC	720

CTCAGTAAAA AACTTGCAAC TCTGGTGACG AACGCACCTG TTGAAGTGGG CTGGGAAGAG 780
 ATGAAATACA GAGGATACGA CAAGAGAAAA CTACTTCCGA TATTGAAAGA ACTGGAGTTT 840
 GCTTCCATCA TGAAGGAACT TCAACTGTAC GAAGAAGCAG AACCCACCGG ATACGAAATC 900
 GTGAAGGATC ATAAGACCTT CGAAGATCTC ATCGAAAAGC TGAAGGAGGT TCCATCTTTT 960
 GCCCTGGACC TTGAAACGTC CTCCTTGGAC CCGTTCAACT GTGAGATAGT CGGCATCTCC 1020
 GTGTGCTTCA AACCGAAAAC AGCTTATTAC ATTCCACTTC ATCAGAGAAA CGCCCACAAT 1080
 CTTGATGAAA CACTGGTGCT GTCGAAGTTG AAAGAGATCC TCGAAGACCC GTCTTCGAAG 1140
 ATTGTGGGTC AGAACCCTGAA GTACGACTAC AAGGTTCTTA TGGTAAAGGG TATATCGCCA 1200
 GTTTATCCGC ATTTTGACAC GATGATAGCT GCATATTGTC TGGAGCCAAA CGAGAAAAAA 1260
 TTCAATCTCG AAGATCTGTC TTTGAAATT CTGGATACA AAATGACGTC 1310

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 436 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Ala Arg Leu Phe Leu Phe Asp Gly Thr Ala Leu Ala Tyr Arg Ala
 1 5 10
 Tyr Tyr Ala Leu Asp Arg Ser Leu Ser Thr Ser Thr Gly Ile Pro Thr
 20 25 30
 Asn Ala Val Tyr Gly Val Ala Arg Met Leu Val Lys Phe Ile Lys Glu
 35 40 45
 His Ile Ile Pro Glu Lys Asp Tyr Ala Ala Val Ala Phe Asp Lys Lys
 50 55 60
 Ala Ala Thr Phe Arg His Lys Leu Leu Val Ser Asp Lys Ala Gln Arg
 65 70 75 80
 Pro Lys Thr Pro Ala Leu Leu Val Gln Gln Leu Pro Tyr Ile Lys Arg
 85 90 95
 Leu Ile Glu Ala Leu Gly Phe Lys Val Leu Glu Leu Glu Gly Tyr Glu
 100 105 110

Ala Asp Asp Ile Ile Ala Thr Leu Ala Ala Lys Gly Cys Thr Phe Phe
115 120 125

Asp Glu Ile Phe Ile Ile Thr Gly Asp Lys Asp Met Leu Gln Leu Val
130 135 140

Asn Glu Lys Ile Lys Val Trp Arg Ile Val Lys Gly Ile Ser Asp Leu
145 150 155 160

Glu Leu Tyr Asp Ser Lys Lys Val Lys Glu Arg Tyr Gly Val Glu Pro
165 170 175

His Gln Ile Pro Asp Leu Leu Ala Leu Thr Gly Asp Asp Ile Asp Asn
180 185 190

Ile Pro Gly Val Thr Gly Ile Gly Glu Lys Thr Ala Val Gln Leu Leu
195 200 205

Gly Lys Tyr Arg Asn Leu Glu Tyr Ile Leu Glu His Ala Arg Glu Leu
210 215 220

Pro Gln Arg Val Arg Lys Ala Leu Leu Arg Asp Arg Glu Val Ala Ile
225 230 235 240

Leu Ser Lys Lys Leu Ala Thr Leu Val Thr Asn Ala Pro Val Glu Val
245 250 255

Asp Trp Glu Glu Met Lys Tyr Arg Gly Tyr Asp Lys Arg Lys Leu Leu
260 265 270

Pro Ile Leu Lys Glu Leu Glu Phe Ala Ser Ile Met Lys Glu Leu Gln
275 280 285

Leu Tyr Glu Glu Ala Glu Pro Thr Gly Tyr Glu Ile Val Lys Asp His
290 295 300

Lys Thr Phe Glu Asp Leu Ile Glu Lys Leu Lys Glu Val Pro Ser Phe
305 310 315 320

Ala Leu Asp Leu Glu Thr Ser Ser Leu Asp Pro Phe Asn Cys Glu Ile
325 330 335

Val Gly Ile Ser Val Ser Phe Lys Pro Lys Thr Ala Tyr Tyr Ile Pro
340 345 350

Leu His His Arg Asn Ala His Asn Leu Asp Glu Thr Leu Val Leu Ser
355 360 365

Lys Leu Lys Glu Ile Leu Glu Asp Pro Ser Ser Lys Ile Val Gly Gln
370 375 380

Asn Leu Lys Tyr Asp Tyr Lys Val Leu Met Val Lys Gly Ile Ser Pro
385 390 395 400

Val Tyr Pro His Phe Asp Thr Met Ile Ala Ala Tyr Leu Leu Glu Pro
 405 410 415
 Asn Glu Lys Lys Phe Asn Leu Glu Asp Leu Ser Leu Lys Phe Leu Gly
 420 425 430
 Tyr Lys Met Thr
 435

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 687 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TCGTACCNGG GNTCNCNANA TCGACTGCAG CATGCAAGCT GGCTAATCAT GGTCATAGCT	60
GTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG CCGGAAGCAT	120
AAAGTGTAAG GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC	180
ACTGCCCCGT TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAAATGAA TCGGCCAACG	240
CGCGGGGAGA GCGGGTTTGC GTATTGGGCG CTCTCCGCT TCCTCGCTCA CTGACTCGCT	300
GCGCTCGGTC GTTCGGCTGC GCGGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT	360
ATCCACAGAA TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC	420
CAGGAACCGT TAAAAAGGCC GCGTTGCTGG GCGTTTTTCC ATAGGCTCCG CCCCCCTTGA	480
CGAGCATCAC AAAAATTCTGA CGCTTCAAGT TCAGAGGTGG GCGAAACCCG ACAGGGACTA	540
TAAAGATTAC CAGGGCGTTT TCCCCCTGGG AAGCTNCCTT CGTGCGCTCT CCTGTTCCCG	600
AACCTGGCCG GTTTAACCGG GATACCNGNT CGGCCTTTTN TCCCCTTNGG GGGAANCCTT	660
GGGGNTTTTN GNAAAANGCT AAGGGTT	687

(2) INFORMATION FOR SEQ ID NO:20:

66710-5-26350

- (A) LENGTH: 701 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCTCGTACCG	GGGATCTNNN	ANATCGACTG	CAGCATGCAA	GCTTGGCGTA	ATCATGGTCA	60
TAGCTGTTTC	CTGTGTGAAA	TTGTTATCCG	CTCACAATTC	CACACAACAT	ACGAGCCGGA	120
AGCATAAAGT	GTAAAGCCTG	GGGTGCCTAA	TGAGTGAGCT	AACTCACATT	AATTGCGTTG	180
CGCTCACTGC	CCGCTTTCCA	GTCGGGAAAC	CTGTCGTGCC	AGCTGCATTA	ATGAATCGGC	240
CAACGCGCGG	GGAGAGGCGG	TTTGCTATT	GGGCGCTCTT	CCGCTTCCTC	GCTCACTGAC	300
TCGCTGCGCT	CGGTCGTTCTG	GCTGCGGCGA	GCGGTATCAG	CTCACTCAAA	GGCGGTAATA	360
CGGTTATCCA	CAGAATCAGG	GGATAACGCA	GGAAAGAACA	TGTGAGCAAA	AGGCCAGCAA	420
AAGGCCAGGA	ACCGTAAAAA	GGCCGCGTTG	CTGGGCGTTT	TTTCCATAGG	CTCCGCCCCC	480
CTGANGAGCA	TCANAAAAAT	CGANGCTCAN	GTCANAGGTG	GCGAAACCCG	ACAGGNETAT	540
TAAAAGATNC	CCAGGCGTTT	TCCCCCCTGG	GAAGCTCCCT	CGTGGGGCTC	TCCTGGTTNC	600
GGNNCCCTGN	CCGGNTTACC	GGGGATAANC	TTGTTCCGGN	CTTTNTCCCC	TTCNNGGAAA	660
ANGGTGGGGG	GTTTTNTNNA	AAAGGCTCAA	AGGCTGGTAN	G		701

(2) INFORMATION/ FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GNTNTAGNNN GGNCTAANNG GCGGGGAAAT CGAGCTCGGT ACCCGGGGAT CCTCTAGAGT 60

CGACCTGCAG GCATGCAAGC TTGGCGTAAT CATGGTCATA GCTGTTTCCT GTGTGAAATT 120
GTTATCCGCT CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT AAAGCCTGGG 180
GTGCCTAATG AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC GCTTTCCAGT 240
CGGGAAACCT GTCGTGCCAG CTGCATTAAT GAATCGGCCA ACGCGCGGGG AGAGGCGGTT 300
TGCGTATTGG GCGCTCTTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCGTTCCGGC 360
TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA GAAATCAGGG 420
GATAACGCAG GGAAAGAACA TGTGAGCAAA AGGCCAGCA AAAGGCCAGG AACCCGTAAA 480
AAGGCCGCGT TGCCTGGCGT TTTTCCATAG GCTCCGCCCC CTTGACGAG CAATCACAAA 540
AATCGACGCT CAAAGTCAAG AGGTGGCGAA ACCCCGACAG GGACTTATAA AGATACCCAG 600
GCCGTTTCCC CCTGGAAGCT CCCCTCCGTG CGCTTCTCCT TGGTTCCCGA CCCTGCCGCT 660
TTACCNGGAT NCCTGTCCGC CCTTTTNTCC CTTTCNGGNA ACCGGGCGCT TTTTTTT 717

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 713 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

NNNNNNNNNG GCTGANAGCG ATAAATCGAG CTCGGTACCC GGGGATCCTC TAGAGTCGAC 60
CTGCAGGCAT GCAAGCTTGG CGTAATCATG GTCATAGCTG TTTCTGTGT GAAATTGTTA 120
TCCGCTCACA ATTGCACACA ACATACGAGC CGGAAGCATA AAGTGTAAG CCTGGGGTGC 180
CTAATGAGTG AGCTAACTCA CATTAAATTGC GTTGCGCTCA CTGCCCCTT TCCAGTCGGG 240
AAACCTGTCTG TGCCAGCTGC ATTAATGAAT CGGCCAACGC GCGGGGAGAG GCGGTTTGCG 300
TATTGGGCGC TCTTCCGCTT CCTCGCTCAC TGA CTGCTGCGT CGCTCGGTG TTCGGCTGCG 360
GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA 420
CGCAGGAAG AACATGTTGA GCAAAGGCC AGCAAAGGC CAGGAACCGT AAAAAGGCCG 480

CGTTTGCTGG CGTTTTTCCC ATAGGCTCCG CCCCCCTTGA CGAACCATCA CAAAAATCGA 540
CGCTCAATTC AGAAGTTGGC GAAAACCCGA CAGGACTAAT AAAGATACCC AGCGTTTCCC 600
CCCCTGGAAA CTCCCCTCCG TTGCGCCTCT CCCTGTTCCC GAACCTTGCC CGCTTACCGG 660
GAATACCTTG TCCNCCTTTT CTCCCCTTCC GGGAANCGTT NGCGCCTTTC CCC 713

CGTTTGCTGG
CGCTCAATTC
CCCCTGGAAA
GAATACCTTG